Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

Communicated by T. M. Sonneborn, August 18, 1977

ABSTRACT **A phylogenetic analysis based upon ribosomal RNA sequence characterization reveals that living systems represent one of three aboriginal lines of descent: (*i*) the eubacteria, comprising all typical bacteria; (*ii*) the archaebacteria, containing methanogenic bacteria; and (*iii*) the urkaryotes, now represented in the cytoplasmic component of eukaryotic cells.**

The biologist has customarily structured his world in terms of certain basic dichotomies. Classically, what was not plant was animal. The discovery that bacteria, which initially had been considered plants, resembled both plants and animals less than plants and animals resembled one another led to a reformulation of the issue in terms of a yet more basic dichotomy, that of eukaryote versus prokaryote. The striking differences between eukaryotic and prokaryotic cells have now been documented in endless molecular detail. As a result, it is generally taken for granted that all extant life must be of these two basic types.

Thus, it appears that the biologist has solved the problem of the primary phylogenetic groupings. However, this is not the case. Dividing the living world into *Prokaryotae* and *Eukaryotae* has served, if anything, to obscure the problem of what extant groupings represent the various primeval branches from the common line of descent. The reason is that eukaryote/prokaryote is not primarily a phylogenetic distinction, although it is generally treated so. The eukaryotic cell is organized in a different and more complex way than is the prokaryote; this probably reflects the former's composite origin as a symbiotic collection of various simpler organisms (1–5). However striking, these organizational dissimilarities do not guarantee that eukaryote and prokaryote represent phylogenetic extremes.

The eukaryotic cell *per se* cannot be directly compared to the prokaryote. The composite nature of the eukaryotic cell makes it necessary that it first be conceptually reduced to its phylogenetically separate components, which arose from ancestors that were noncomposite and so individually are comparable to prokaryotes. In other words, the question of the primary phylogenetic groupings must be formulated solely in terms of relationships among "prokaryotes"—i.e., noncomposite entities. (Note that in this context there is no suggestion *a priori* that the living world is structured in a dichotomous way.)

The organizational differences between prokaryote and eukaryote and the composite nature of the latter indicate an important property of the evolutionary process: Evolution seems to progress in a "quantized" fashion. One level or domain of organization gives rise ultimately to a higher (more complex) one. What "prokaryote" and "eukaryote" actually represent are two such domains. Thus, although it is useful to define phylogenetic patterns within each domain, it is not meaningful to construct phylogenetic classifications between domains: Prokaryotic kingdoms are not comparable to eukaryotic ones. This should be recognized by an appropriate terminology. The highest phylogenetic unit in the prokaryotic domain we think should be called an "urkingdom"—or perhaps "primary kingdom." This would recognize the qualitative distinction between prokaryotic and eukaryotic kingdoms and emphasize that the former have primary evolutionary status.

The passage from one domain to a higher one then becomes a central problem. Initially one would like to know whether this is a frequent or a rare (unique) evolutionary event. It is traditionally assumed—without evidence—that the eukaryotic domain has arisen but once; all extant eukaryotes stem from a common ancestor, itself eukaryotic (2). A similar prejudice holds for the prokaryotic domain (2). [We elsewhere argue (6) that a hypothetical domain of lower complexity, that of "progenotes," may have preceded and given rise to the prokaryotes.] The present communication is a discussion of recent findings that relate to the urkingdom structure of the prokaryotic domain and the question of its unique as opposed to multiple origin.

Phylogenetic relationships cannot be reliably established in terms of noncomparable properties (7). A comparative approach that can measure degree of difference in comparable structures is required. An organism's genome seems to be the ultimate record of its evolutionary history (8). Thus, comparative analysis of molecular sequences has become a powerful approach to determining evolutionary relationships (9, 10).

To determine relationships covering the entire spectrum of extant living systems, one optimally needs a molecule of appropriately broad distribution. None of the readily characterized proteins fits this requirement. However, ribosomal RNA does. It is a component of all self-replicating systems; it is readily isolated; and its sequence changes but slowly with time—permitting the detection of relatedness among very distant species (11–13). To date, the primary structure of the 16S (18S) ribosomal RNA has been characterized in a moderately large and varied collection of organisms and organelles, and the general phylogenetic structure of the prokaryotic domain is beginning to emerge.

A comparative analysis of these data, summarized in Table 1, shows that the organisms clearly cluster into several primary kingdoms. The first of these contains all of the typical bacteria so far characterized, including the genera *Acetobacterium, Acinetobacter, Acholeplasma, Aeromonas, Alcaligenes, Anacystis, Aphanocapsa, Bacillus, Bdellovibrio, Chlorobium, Chromatium, Clostridium, Corynebacterium, Escherichia, Eubacterium, Lactobacillus, Leptospira, Micrococcus, Mycoplasma, Paracoccus, Photobacterium, Propionibacterium,*

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

* Present address: Department of Biophysical Sciences, University of Houston, Houston, TX 77004.

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Pseudomonas, Rhodopseudomonas, Rhodospirillum, Spirochaeta, Spiroplasma, Streptococcus, and *Vibrio* (refs. 13–17; unpublished data). The group has three major subdivisions, the blue-green bacteria and chloroplasts, the "Gram-positive" bacteria, and a broad "Gram-negative" subdivision (refs. 3, 4, 13–17; unpublished data). It is appropriate to call this urkingdom the *eubacteria*.

A second group is defined by the 18S rRNAs of the eukaryotic cytoplasm—animal, plant, fungal, and slime mold (unpublished data). It is uncertain what ancestral organism in the symbiosis that produced the eukaryotic cell this RNA represents. If there had been an "engulfing species" (1) in relation to which all the other organisms were endosymbionts, then it seems likely that 18S rRNA represents that species. This hypothetical group of organisms, in one sense the major ancestors of eukaryotic cells, might appropriately be called *urkaryotes*. Detailed study of anaerobic amoebae and the like (18), which seem not to contain mitochondria and in general are cytologically simpler than customary examples of eukaryotes, might help to resolve this question.

Eubacteria and urkaryotes correspond approximately to the conventional categories "prokaryote" and "eukaryote" when they are used in a phylogenetic sense. However, they do not constitute a dichotomy; they do not collectively exhaust the class of living systems. There exists a third kingdom which, to date, is represented solely by the methanogenic bacteria, a relatively unknown class of anaerobes that possess a unique metabolism based on the reduction of carbon dioxide to methane (19–21). *These "bacteria" appear to be no more related to typical bacteria than they are to eukaryotic cytoplasms.* Although the two divisions of this kingdom appear as remote from one another as blue-green algae are from other eubacteria, they nevertheless correspond to the *same* biochemical phenotype. The apparent antiquity of the methanogenic phenotype plus the fact that it seems well suited to the type of environment presumed to exist on earth 3–4 billion years ago lead us tentatively to name this urkingdom the *archaebacteria*. Whether or not other biochemically distinct phenotypes exist in this kingdom is clearly an important question upon which may turn our concept of the nature and ancestry of the first prokaryotes.

Table 1 shows the three urkingdoms to be equidistant from one another. Because the distances measured are actually proportional to numbers of mutations and not necessarily to time, it cannot be proven that the three lines of descent branched from the common ancestral line at about the same time. One of the three may represent a far earlier bifurcation than the other two, making there in effect only two urkingdoms. Of the three possible unequal branching patterns the case for which the initial bifurcation defines urkaryotes vs. all bacteria requires further comment because, as we have seen, there is a predilection to accept such a dichotomy.

The phenotype of the methanogens, although ostensibly "bacterial," on close scrutiny gives no indication of a specific phylogenetic resemblance to the eubacteria. For example, methanogens do have cell walls, but these do not contain peptidoglycan (24). The biochemistry of methane formation appears to involve totally unique coenzymes (23, 25, 26). The methanogen rRNAs are comparable in size to their eubacterial counterparts, but resemble the latter specifically in neither sequence (Table 1) nor in their pattern of base modification (23). The tRNAs from eubacteria and eukaryotes are characterized by a common modified sequence, TΨCG; methanogens modify this tRNA sequence in a quite different and unique way (23). It must be recognized that very little is known of the general biochemistry of the methanogens—and almost nothing is known regarding their molecular biology. Hence, although the above points are few in number, they represent most of what is now known. There is no reason at present to consider methanogens as any closer to eubacteria than to the "cytoplasmic component" of the eukaryote. Both in terms of rRNA sequence measurement and in terms of general phenotypic differences, then, the three groupings appear to be distinct urkingdoms.

If a third urkingdom exists, does this suggest that many more such will be found among yet to be characterized organisms? We think not, although the matter clearly requires an exhaustive search. As seen above, the number of species that can be classified as eubacteria is moderately large. To this list can be added *Spirillum* and *Desulfovibrio*, whose rRNAs appear typically eubacterial by nucleic acid hybridization measurements (27). Because the list is also phenotypically diverse, it seems unlikely that many, if any, of the yet uncharacterized

prokaryotic groups will be shown to have coequal status with the present three. Conceivably the halophiles whose cell walls contain no peptidoglycan, are candidates for this distinction (28, 29).

Eukaryotic organelles, however, could be a different matter. There can be no doubt that the chloroplast is of specific eubacterial origin (3, 4). A question arises with the remaining organelles and structures. Mitochondria, for example, do not conform well to a "typically prokaryotic" phenotype, which has led some to conclude that they could not have arisen as endosymbionts (30). By using "prokaryote" in a phylogenetic sense, this formulation of the issue does not recognize a third alternative—that the organelle in question arose endosymbiotically from a separate line of descent whose phenotype is not "typically prokaryotic" (i.e., eubacterial). It is thus conceivable that some endosymbiotically formed structures represent still other major phylogenetic groups; some could even be the only extant representation thereof.

The question that remains to be answered is whether the common ancestor of all three major lines of descent was itself a prokaryote. If not, each urkingdom represents an independent evolution of the prokaryotic level of organization. Obviously, much more needs to be known about the general properties of all the urkingdoms before this matter can be definitely settled. At present we can point to two arguments suggesting that each urkingdom does represent a separate evolution of the prokaryotic level of organization.

The first argument concerns the stability of the general phenotypes. The general eubacterial phenotype has been stable for at least 3 billion years—i.e., the apparent age of blue-green algae (31). The methanogenic phenotype seems to be at least this old in that branchings within the two urkingdoms are comparably deep (see Table 1). The time available to form each phenotype (from their common ancestor) is then short by comparison, which seems paradoxical in that the two phenotypes are so fundamentally different. We think that this ostensible paradox implies that the common ancestor in this case was not a prokaryote. It was a far simpler entity; it probably did not evolve at the "slow" rate characteristic of prokaryotes; it did not possess many of the features possessed by prokaryotes, and so these evolved independently and differently in separate lines of descent.

The second argument concerns the quality of the differences in the three general phenotypes. It seems highly unlikely, for example, that differences in general patterns of base modification in rRNAs and tRNAs are related to the niches that organisms occupy. Rather, differences of this nature imply independent evolution of the properties in question. It has been argued elsewhere that features such as RNA base modification generally represent the final stage in the evolution of translation (32). If these features have evolved separately in two lines of descent, their common ancestor, lacking them, had a more rudimentary version of the translation mechanism and consequently, could not have been as complex as a prokaryote (6).

With the identification and characterization of the urkingdoms we are for the first time beginning to see the overall phylogenetic structure of the living world. It is not structured in a bipartite way along the lines of the organizationally dissimilar prokaryote and eukaryote. Rather, it is (at least) tripartite, comprising (*i*) the typical bacteria, (*ii*) the line of descent manifested in eukaryotic cytoplasms, and (*iii*) a little explored grouping, represented so far only by methanogenic bacteria.

The ideas expressed herein stem from research supported by the National Aeronautics and Space Administration and the National Science Foundation. We are grateful to a number of colleagues who have helped to generate the yet unpublished data that make these speculations possible: William Balch, Richard Blakemore, Linda Bonen, Tristan Dyer, Jane Gibson, Ramesh Gupta, Robert Hespell, Bobby Joe Lewis, Kenneth Luehrsen, Linda Magrum, Jack Maniloff, Norman Pace, Mitchel Sogin, Stephan Sogin, David Stahl, Ralph Tanner, Thomas Walker, Ralph Wolfe, and Lawrence Zablen. We thank Linda Magrum and David Nanney for suggesting the name "archaebacteria."

1. Stanier, R. Y. (1970) *Symp. Soc. Gen. Microbiol.* **20**, 1–38.
2. Margulis, L. (1970) *Origin of Eucaryotic Cells* (Yale University Press, New Haven).
3. Zablen, L. B., Kissel, M. S., Woese, C. R. & Buetow, D. E. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2418–2422.
4. Bonen, L. & Doolittle, W. F. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2310–2314.
5. Bonen, L., Cunningham, R. S., Gray, M. W. & Doolittle, W. F. (1977) *Nucleic Acid Res.* **4**, 663–671.
6. Woese, C. R. & Fox, G. E. (1977) *J. Mol. Evol.*, in press.
7. Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical Taxonomy* (W. H. Freeman, San Francisco).
8. Zuckerkandl, E. & Pauling, L. (1965) *J. Theor. Biol.* **8**, 357–366.
9. Fitch, W. M. & Margoliash, E. (1967) *Science* **155**, 279–284.
10. Fitch, W. M. (1976) *J. Mol. Evol.* **8**, 13–40.
11. Sogin, S. J., Sogin, M. L. & Woese, C. R. (1972) *J. Mol. Evol.* **1**, 173–184.
12. Woese, C. R., Fox, G. E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, B. J. & Stahl, D. (1975) *Nature* **254**, 83–86.
13. Fox, G. E., Pechman, K. R. & Woese, C. R. (1977) *Int. J. Syst. Bacteriol.* **27**, 44–57.
14. Uchida, T., Bonen, L., Schaup, H. W., Lewis, B. J., Zablen, L. B. & Woese, C. R. (1974) *J. Mol. Evol.* **3**, 63–77.
15. Zablen, L. B. & Woese, C. R. (1975) *J. Mol. Evol.* **5**, 25–34.
16. Doolittle, W. F., Woese, C. R., Sogin, M. L., Bonen, L. & Stahl, D. (1975) *J. Mol. Evol.* **4**, 307–315.
17. Pechman, K. J., Lewis, B. J. & Woese, C. R. (1976) *Int. J. Syst. Bacteriol.* **26**, 305–310.
18. Bovee, E. C. & Jahn, T. L. (1973) in *The Biology of Amoeba*, ed. Jeon, K. W. (Academic Press, New York), p. 38.
19. Wolfe, R. S. (1972) *Adv. Microbiol. Phys.* **6**, 107–146.
20. Zeikus, J. G. (1977) *Bacteriol. Rev.* **41**, 514–541.
21. Zeikus, J. G. & Bowen, V. G. (1975) *Can. J. Microbiol.* **21**, 121–129.
22. Balch, W. E., Magrum, L. J., Fox, G. E., Wolfe, R. S. & Woese, C. R. (1977) *J. Mol. Evol.*, in press.
23. Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S. & Woese, C. R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 4537–4541.
24. Kandler, O. & Hippe, H. (1977) *Arch. Microbiol.* **113**, 57–60.
25. Taylor, C. D. & Wolfe, R. S. (1974) *J. Biol. Chem.* **249**, 4879–4885.
26. Cheeseman, P., Toms-Wood, A. & Wolfe, R. S. (1972) *J. Bacteriol.* **112**, 527–531.
27. Pace, B. & Campbell, L. L. (1971) *J. Bacteriol.* **107**, 543–547.
28. Brown, A. D. & Cho, K. Y. (1970) *J. Gen. Microbiol.* **62**, 267–270.
29. Reistad, R. (1972) *Arch. Mikrobiol.* **82**, 24–30.
30. Raff, R. A. & Mahler, H. R. (1973) *Science* **180**, 517–521.
31. Shopf, J. W. (1972) *Exobiology—Frontiers of Biology* (North Holland, Amsterdam), Vol. 23, pp. 16–61.
32. Woese, C. R. (1970) *Symp. Soc. Gen. Microbiol.* **20**, 39–54.